International Journal of Serious Games

ISSN: 2384-8766
https://journal.seriousgamessociety.org/

Article

GEMS: Evaluating the Relative Importance of Game Attributes

Ernesto Pacheco-Velazquez[1], Alejandro Flores-Benitez[2], Astrid Rodriguez-Castelblanco[2], Ivan Arana-Solares[2], Sergio Ramirez-Echeverri[3]

[1]Institute for the Future of Education, Tecnologico de Monterrey, Monterrey, Mexico; [2]School of Engineering and Sciences, Tecnologico de Monterrey, Ciudad de Mexico, Mexico; [3]Department of Production Engineering, Universidad EAFIT, Medellin, Colombia

epacheco@tec.mx; alejandro.flores.benitez@tec.mx; astrid.rodriguez@tec.mx; iarana@tec.mx; sramire@edu.eafit.co

Keywords:

Serious Games
Game Evaluation
User Experience
Multidimensional Scale
Game Design
Player Perception
Video Game Assessment

Received: October 2025
Accepted: December 2025
Published: January 2026
DOI: 10.17083/h569ps87

Abstract

This study presents the Game Experience Multidimensional Scale (GEMS), a 13-item instrument designed to identify the attributes that players consider most influential when evaluating video games. Unlike existing scales that measure the presence of certain attributes or their intensity, GEMS focuses on the relative importance that users assign to each attribute when forming an overall opinion of a game. The scale was developed through an extensive review of the literature on existing evaluation models and refined with expert input. It was then applied to a sample of 619 students, primarily at the high school and undergraduate levels, who evaluated a game of their choice using the GEMS questionnaire. The instrument demonstrated high internal consistency (Cronbach's $\alpha = 0.861$). Two analytical methods (stepwise multiple linear regression and linear programming) were employed to determine the attributes that best explain users' overall ratings, with the regression model explaining 48.99% of the variance in overall game evaluation. Although both methods converged on similar key predictors, the findings should be interpreted with caution, as the homogeneity of the sample, the genre of the evaluated games, or demographic factors may have influenced the results. Nevertheless, GEMS offers a novel and preliminary approach to game evaluation, with potential applications in educational and professional contexts once further validation is conducted.

1. Introduction

Recent data on video games show that their use has experienced exponential growth and that they have become one of the most widespread, profitable, and influential forms of entertainment in the modern world [1]. Moreover, video games have developed into a powerful

industry and an established reality. Some of the most popular games attract millions of users [2]. The cultural phenomenon has become so significant that today there are videos and live streams where millions of people gather to watch others play different types of video games. These gatherings represent new forms of community and digital socialization, making such events particularly popular among young people [3], and reinforcing the identity of the gamer as an emerging social category [4]. Watching others play is often associated with experiences of learning, aspiration, and emotional engagement, contributing to a growing sense of belonging to a community [5]. The way participants interact within these communities is often astonishing to those unfamiliar with such events. In these spaces, users regularly share opinions and experiences related to the game. Comments range from first-time visitors to seasoned players expressing their fascination with these events [2]. Participants also discuss various characteristics of the games, such as the level of challenge, emotional engagement, story narrative, animation, and other key features. This level of emotional involvement, along with the richness of interactive elements that characterize video games, has drawn the attention of researchers from multiple disciplines. In particular, the educational field has begun to explore how these environments can be used as training tools, giving rise to a new category: serious games.

Over the past two decades, video games have not only been viewed as entertainment but have also become the subject of numerous multidisciplinary scientific studies concerning their applications in various contexts. The importance of video games in society has reached such magnitude that measuring user satisfaction has become essential [6]. As a result, several scales have been developed to capture the user experience and satisfaction during interaction with digital games. Each scale aims to systematically capture the factors or attributes that influence player engagement.

The use of games in different contexts has made their evaluation a key area of research. Assessing these games allows developers and educators to verify whether entertainment, learning, or competency development objectives are being met [7]. The evaluation of methodologies or scales used to measure the effectiveness of serious games remains an area with multiple opportunities for development [8]. Some studies highlight the lack of uniform evaluation standards, while others suggest the need to develop tools adapted to the specific context in which the games are used. Several evaluation scales have been developed, such as the Game Engagement Questionnaire (GEQ) [9], eGameFlow [7], and the Serious Game Quality Model (SGQM) [10] among others. However, there is no consensus on the optimal approach for evaluating a game or on which elements users consider most relevant in an evaluation [6]. Despite growing interest, systematic tools for evaluating the quality and impact of these games remain limited [11].

Based on the above and considering the need for a tool that allows for the systematic evaluation of video games, categorization of their key features, and identification of opportunities for improvement, this article proposes the Game Experience Multidimensional Scale (GEMS). This 13-item scale was developed through a rigorous and comprehensive review of existing literature on game evaluation scales, synthesizing the most relevant attributes identified across multiple studies. Its construction further integrated insights from both user perceptions and expert validation from professionals in game design and interactive experience assessment, ensuring that the scale reflects both theorical foundations and practical relevance. The aim of GEMS is to accurately capture the attributes players consider most influential in their overall judgment of a video game, offering a structured, adaptable, and evidence-based instrument suitable for academic, professional, and research applications.

Although GEMS was conceived with potential applicability to both entertainment and serious games, the present study primarily draws on data from entertainment titles. Therefore, its current contribution should be understood as a first step toward developing an instrument that can later be adapted and validated in educational, professional, or therapeutic contexts.

2. Research Objectives

The objectives of this article are to:

1. Propose a scale that consolidates a wide range of elements suggested by various authors into thirteen distinct attributes.
2. Validate the proposed scale.
3. Identify, through two complementary methodologies, the factors that have the greatest influence on user satisfaction when using a game.

3. Literature Review

The transformation that video games have undergone since their emergence in the second half of the 20th century has been radical. Their complexity spans various aspects, including the technology used, story narrative, character design, multiplayer game development, internet integration, and the implementation of elements such as virtual and augmented reality, among others [12]. The consumption of these products has become so significant that the variety of genres now aims to reach all types of audiences and pursue objectives that go beyond mere entertainment [13].

An important classification distinguishes between games designed exclusively for entertainment and those with purposes beyond recreational experience, which are defined as Serious Games (SG). While both types of games share similar mechanics and dynamics [14], SGs are used in educational contexts such as learning, training, behavior change, or awareness raising, among others [15]. Their applications also extend to other domains, including therapy and health promotion, where games are leveraged to support rehabilitation or cognitive training. The line between commercial games and serious games is so subtle that some authors argue that it lies not in the mechanics or design, but in the intent and context of use. Abt coined the term and noted that any game can be considered serious if its use pursues objectives that go beyond entertainment [16]. This definition has gained strength with recent research and a growing literature that recognizes the educational value of video games [17] . Therefore, evaluating user satisfaction in video games is relevant even in non-commercial contexts. The need to measure the impact of these tools in terms of learning, knowledge acquisition, skill development, or behavioral change is critical in different settings, so establishing theoretical frameworks that capture key attributes is essential.

The effectiveness of a video game depends on a broad set of attributes ranging from technical features such as animation and sound to elements that involve players' emotional engagement. For example, eGameFlow identifies eight attributes (Concentration, Goal Clarity, Feedback, Challenge, Autonomy, Immersion, Social Interaction, and Knowledge Improvement), which are based on the flow theory applied to game design. This scale is focused on measuring how absorbing and rewarding the player's experience is. On the other hand, the Serious Game Quality Model (SGQM) aims to assess the quality of serious games from

technical, pedagogical, and functional perspectives. Its items are grouped into three main categories: the definition of objectives, product quality, and effectiveness and impact.

Table 1 presents a set of scales, the approach or intended purpose behind their development, the attributes each scale proposes for evaluating a video game, and the bibliographic references from which these scales were obtained. As can be seen, there are various approaches from which games can be evaluated, for example, flow, the technical quality of the game, motivation, degree of immersion, among others. Most scales typically focus on a particular perspective. There are few references in which attributes are evaluated with the aim of measuring the overall user experience in serious games. For instance, the authors of references [18] and [19] assess attributes related to gaming experience, learning experience, adaptivity, usability, and fidelity.

Table 1. Description of the scales and the attributes evaluated

Name of the Scale	Acronym	Approach	Evaluated Attributes	References
EGameFlow	EGF	Player experience / flow (serious games)	Concentration, Goal Clarity, Feedback, Challenge, Autonomy, Immersion, Social Interaction, Knowledge Improvement	[7]
Game Experience Questionnaire	GEQ	Emotional/cognitive experience	Immersion, Flow, Competence, Positive/Negative Affect, Tension, Challenge; Social and Post-Game modules	[9]
Serious Game Quality Model	SGQM	Technical-pedagogical quality (serious games)	Functionality, reliability, usability, efficiency, maintainability, portability, content, feedback, rewards, effectiveness, appeal, and balance	[10]
Game User Experience Satisfaction Scale	GUESS	Player Satisfaction	Usability/Playability, Narrative, Engrossment, Enjoyment, Creative freedom, Audio/Visual aesthetics, Personal gratification, Social connectivity	[20]
Immersive Experience Questionnaire	IEQ	Degree of immersion	Control, challenge, dissociation from reality, concentration, curiosity	[21]
Presence Questionnaire	PQ	Presence in virtual environments	Sensory attention, emotional involvement, realism/presence	[22]
Flow State Scale / Dispositional Flow Scale	FSS/DFS	General flow state	Goal Clarity, concentration, control, Feedback, action-awareness fusion, time distortion, loss of self-consciousness, enjoy	[23] [24]
Player Experience of Need Satisfaction	PENS	Motivation / self-determination	Competence, autonomy, social relatedness, intuitive controls, immersion	[25]
Intrinsic Motivation Inventory	IMI	Intrinsic motivation	Interest/enjoyment, effort, perceived value, pressure/tension, competence, autonomy	[26]
Hexad User Types Scale	Hexad	Motivational typology (gamification)	Socialiser, Free Spirit, Achiever, Philanthropist, Player, Disruptor	[27]
Enjoyment Questionnaire	SoEQ	Origins of enjoyment	Humor, relaxation, savoring, variety, purpose, social responsibility	[28]

Serious Games Evaluation Scale	SGES	General evaluation of serious games	Presence, enjoyment, perceived learning, narrative, realism, audiovisual feedback, relevance, goal clarity, usability, learning, motivation	[29]
Gameplay Scale for Educative Games	-	Evaluation in children	Usefulness, usability, goal clarity, concentration, enjoyment	[30]
Gaming Educational Balanced Model	GEB	Game/play and educational balance	Balance between recreational experience and learning, type of learning (Can Learn vs Must Learn)	[31]
Game-Based Learning Evaluation Model	GEM	Methodology for educational games	Action language (usability), feedback, challenge, control, effective learning	[32]
Playability Model: Characterization of the Player Experience	-	Evaluation of player experience (playability as quality of use)	Satisfaction, Learning, Effectiveness, Immersion, Motivation, Emotion, Socialization. Includes sub-attributes like difficulty, discovery, concentration, realism, curiosity, stimulation, sensory appeal, communication, etc. Also proposes facets and a heuristic evaluation tool (PHET)	[33] [34]

However, even when there are scales that assess more than one single perspective, there is no tool that allows us to determine which of these attributes users consider key in forming their overall opinion about the game. Evaluation methodologies that measure users' overall perception of different types of games remain an underexplored area. Some studies highlight the absence of consistent evaluation patterns and acknowledge the need to develop tools that are adapted to the context or the specific game genre [8]. Other authors point out the lack of consensus on the most appropriate approach to comprehensively assess the various aspects of a serious game [6]. In summary, although we observe exponential growth in the use of serious games across different contexts, there is still a lack of systematic tools to assess the overall quality of these resources, identify which attributes are key to maximizing user satisfaction, and evaluate their impact on players depending on the game genre [11].

On the other hand, while many of the evaluation instruments traditionally used in video game research (e.g., GEQ, eGameFlow, GUESS, PENS) originate from earlier years, more recent work has begun to address the unique demands of serious games and educational contexts. For example, Carrión-Toro et al. propose a usability instrument tailored for serious games in learning settings [35]. The SEGiNAS scale was developed specifically for serious educational games in nursing, capturing dimensions such as engagement, teaching effectiveness, practical application and content relevance [36]. A stakeholder-centered framework offers a holistic approach to serious game design and evaluation, including expert review and metrics relevant for educational outcomes [37]. Additionally, systematic reviews on serious games for mental health literacy highlight the need for rigorous evaluation of user experience, effectiveness, adherence, and in-game performance alongside self-reported measures [38]. These recent contributions offer helpful models but also confirm the scarcity of instruments that simultaneously prioritize user perceptions of attribute importance, statistical rigor, and applicability across domains.

4. Methodology

This study was conducted with the aim of constructing an evaluation instrument capable of identifying the attributes users consider most relevant in assessing their overall experience when interacting with a serious video game. To this end, a structured methodological strategy was followed, consisting of three stages: initial attribute selection, user-based validation, and expert validation.

4.1 Creation of the scale for the global evaluation of a video game

The development of a global evaluation scale for video games was carried out through the following steps:

1. Initial Selection of Attributes: In the first stage, an exhaustive review of the specialized literature on scales used to evaluate video games was conducted, including instruments such as EGameFlow [7], GEQ [9], SGQM [10], GUESS [20], PENS [25], among others. Based on this review, 20 key and consistently recurring attributes in the evaluation of digital games were identified and systematically defined, encompassing technical, cognitive, motivational, and usability dimensions. Attributes were regarded as consistently recurring when similar constructs emerged across different evaluation models, even if labeled differently or embedded within broader dimensions. For instance, feedback-related aspects appeared under various formulations across player experience–oriented and pedagogical evaluation frameworks, indicating conceptual convergence rather than simple frequency of mention.

2. User Consultation: To incorporate the end-user perspective, a survey was designed and administered to a purposive sample of 217 university students, who responded anonymously via a digital form. Participants were asked to select the ten attributes they considered most important when evaluating a video game based on their personal experience. The results of this survey helped identify the most frequently selected attributes. At this stage, participants were not asked about their perceived relevance of the selected factors; they were simply asked to mention the ten attributes they considered most important. The results were ranked according to the number of mentions. While several attributes received similar scores, the differences between positions 8 and 12 were minimal (with a maximum difference of three mentions between adjacent attributes), a clearer gap emerged after position 12, where the number of mentions dropped by 17, representing a markedly larger decrease than those observed among the preceding ranks. Additionally, the difference between positions 13 and 14 was even more pronounced, with a decrease of 24 mentions.

3. Expert Validation and Final Instrument Design: The results obtained were reviewed and evaluated by an expert panel composed of eight professionals with relevant experience: four university professors specializing in educational technology, two video game designers, one programmer, and one expert in entertainment software marketing. The panel indicated that a 20-item instrument would be excessively long for practical application and recommended reducing the number of attributes to be evaluated by users. Given the difference in the number of mentions between the attribute ranked 12th and the next highest, it was suggested to construct a questionnaire around the twelve highest-scoring attributes. Following the recommendation of one expert, an additional item was incorporated, resulting in a final 13-item instrument,

which constitutes the proposed GEMS. The experts also suggested using a 10-point Likert scale to capture sufficient variability in responses. Furthermore, the wording of the items was modified for clarity. To minimize response bias, the order of the items was randomized in the final survey.

The use of expert judgment to refine, merge, or eliminate items is aligned with established practices in psychometric test development, where expert review is a key step for ensuring content validity and conceptual coherence before large-scale validation [39] [40].

The final version of the questionnaire, which we have named GEMS (Game Experience Multidimensional Scale), is presented in Table 2. It is important to note that the section of the survey collecting general demographic information from participants (such as age, gender, and educational level) has been omitted from the table for brevity. Additionally, participants were asked not only to evaluate a specific game, but also to indicate the genre of the game they were assessing (e.g., action, adventure, sports, puzzle, strategy, etc.).

Participants were first asked to provide the name of the video game to be evaluated, in order to categorize the game appropriately. For each of the 13 items of the GEMS scale, two questions were posed: one regarding the perceived importance of the attribute in video games in general (importance), and another assessing how the participant rated that specific attribute in the game under evaluation (game rating). Finally, participants provided an overall assessment of the video game. This resulted in a questionnaire consisting of 28 core questions, in addition to items collecting demographic information.

Table 2. Final version of the questionnaire

From this point on, we will ask you to evaluate the importance of various game attributes. You will also be asked to rate a specific game. For each attribute, use a scale from 1 to 10:

- **Importance: How important is this attribute in a game, based on your profile as a player? (1 = Not important at all, 10 = Extremely important)**
- **Game Rating: How would you rate the game you're evaluating in terms of this attribute? (1 = Terrible, 10 = Excellent)**

In your experience, how important is the following attribute in a game?

Question	Item	Answer
1	**Name of the game you are evaluating** (preferably, choose a game you are very familiar with):	Text
2	**Challenge**: The game's difficulty level should evolve in line with your skill progression, keeping it interesting and recreational enjoyment without becoming frustrating.	[1 – 10]
3	**Challenge (Game Rating)**: How well does the game adjust its difficulty level to match your skill progression, keeping the experience engaging and rewarding without becoming frustrating?	[1 – 10]
4	**Feedback**: Consistent feedback helps you assess your actions, improve, and feel a sense of accomplishment as you progress.	[1 – 10]
5	**Feedback (Game Rating)**: How well does this game provide you with consistent feedback? Think about whether it helps you understand your actions, improve, and feel satisfied with your progress.	[1 – 10]
6	**Immersion**: Immersion occurs when the game fully engages you, making it enjoyable and encouraging longer play sessions.	[1 – 10]
7	**Immersion (Game Rating)**: How well does the game immerse you in its world? Does it make you feel completely absorbed and deeply engaged?	[1 – 10]

8	**Concentration**: Concentration is key to fully enjoying a game, helping you stay focused without unnecessary distractions.	[1 – 10]
9	**Concentration (Game Rating)**: How well does the game keep your attention? Does it keep you focused and in "game mode" without unnecessary interruptions?	[1 – 10]
10	**Clarity of Objectives**: How important is it for a game to have clear goals? When both main and secondary goals are easy to understand, it's easier to enjoy the game without getting lost.	[1 – 10]
11	**Clarity of Objectives (Game Rating)**: How clear are the goals in this game? Do they help you easily understand what you need to do and play without confusion?	[1 – 10]
12	**Autonomy**: How important is it for a game to give you autonomy? Autonomy means being able to make your own decisions and feeling in control, creating a personalized gameplay experience.	[1 – 10]
13	**Autonomy (Game Rating):** How well does the game allow you to feel in control of your decisions and actions, supporting a personalized experience?	[1 – 10]
14	**Social Interaction**: Social interaction lets you collaborate or compete with others, enhancing player enjoyment and fostering a sense of community.	[1 – 10]
15	**Social Interaction (Game Rating)**: How well does the game allow you to interact with other players? Does it offer good opportunities for collaboration or competition and support social enjoyment?	[1 – 10]
16	**Knowledge Enhancement**: How important is it for a game to help you learn something new? Knowledge enhancement occurs when the game teaches something useful or helps you develop skills beyond entertainment.	[1 – 10]
17	**Knowledge Enhancement (Game Rating):** How well do you think this game helps you learn or improve skills? Do you feel it has educational value in addition to its entertainment value?	[1 – 10]
18	**Emotional Involvement:** How important is it for a game to make you feel emotions? Emotional involvement happens when the game moves you, whether through its story, characters, or unique moments.	[1 – 10]
19	**Emotional Involvement (Game Rating):** How well does this game move you emotionally? Does it engage you through achievements, storytelling, characters, or special moments?	[1 – 10]
20	**Balance between Skills and Tasks:** How important is it for a game to balance your skills with the challenges it presents? A good balance makes the game exciting and stimulating without being too hard or too easy.	[1 – 10]
21	**Balance (Game Rating):** How well does the game balance your skills with its challenges? Does it keep you motivated and entertained without being too frustrating or boring?	[1 – 10]
22	**Engaging Narrative**: How important is it for a game to have a compelling story? An engaging narrative makes the game more exciting, gives meaning to your actions, and creates a memorable experience.	[1 – 10]
23	**Engaging Narrative (Game Rating):** How interesting and immersive is the game's story? Does it draw you in and give meaning to your in-game actions?	[1 – 10]
24	**Progression Structure:** How important is it for a game to recognize your progress? A good progression system motivates you by rewarding achievements and encouraging continued gameplay.	[1 – 10]
25	**Progression Structure (Game Rating):**	[1 – 10]

	How well does this game acknowledge your progress? Does it motivate you through rewards or incentives as you achieve things?	
26	**Animation and Sound:** How important is it for a game to have high-quality animation and sound? The quality of animation and sound enhances realism and immersion, making the game more visually and auditorily engaging.	[1 – 10]
27	**Animation and Sound (Game Rating):** How well do the animations and sound design enhance the experience? Do they make the game feel smooth, visually appealing, and auditorily immersive?	[1 – 10]
28	**Overall Game Rating (Game Rating):** Overall, how would you rate this game? Consider all the aspects that make the game enjoyable, exciting, and memorable.	[1 – 10]

It is important to clarify that user satisfaction in this study was not conceived as the cumulative outcome of a game meeting all the attributes. Instead, it was measured through an overall game rating (Question 28), which represents the participant's independent and holistic opinion of the game. This overall score was included in the instrument as a dependent variable for multiple regression and linear programming analyses, and the inclusion of this question is what allows the relative importance of each attribute to be estimated.

4.2 Data analysis methods

The survey was administered to a total of 658 participants, primarily composed of high school students and university students. After excluding individuals who either declined to provide consent for research use or indicated that they did not play video games, the final sample consisted of 619 valid responses.

To assess the internal consistency of the instrument, Cronbach's alpha was calculated across the 13 items. The resulting value of $\alpha = 0.861$ indicates a high level of internal reliability, suggesting that the items in the scale are coherently measuring a common construct related to user perceptions of important video game attributes.

Table 3 presents the descriptive statistics (mean and standard deviation) for each of the 13 attributes evaluated, as well as the corrected item–total correlation and Cronbach's alpha if item deleted. These indicators help assess the contribution of each item to the overall reliability of the scale.

Table 3. Descriptive statistics and reliability indicators for each attribute

Attribute	Mean	St. Dev.	Alpha if Item is deleted	Corrected Item-Total Correlation
Challenge	8.50	1.86	0.8512	0.5240
Feedback	7.82	2.35	0.8563	0.4653
Immersion	8.97	1.53	0.8505	0.5547
Concentration	8.58	1.73	0.8515	0.5228
Clarity of Goals	8.69	1.78	0.8501	0.5450
Autonomy	8.67	1.79	0.8505	0.5372
Social Interaction	8.01	2.34	0.8615	0.3916
Knowledge Enhancement	7.69	2.41	0.8539	0.5032
Emotional involvement	8.30	2.07	0.8481	0.5718
Balance between skills and task	8.69	1.69	0.8433	0.6770
Engaging Narrative	8.43	2.16	0.8504	0.5395
Progression structure	8.77	1.70	0.8433	0.6760
Animation and sound	8.99	1.55	0.8542	0.4752
Total	110.13	15.46		

Table 3 shows that the attributes with the highest mean scores were Animation and Sound (M = 8.99) and Immersion (M = 8.97), indicating that participants perceive these elements as particularly important in video games. In contrast, Knowledge Enhancement (M = 7.69) and Feedback (M = 7.82) received the lowest mean scores, although still relatively high on a 10-point scale. It is important to clarify at this point that students were not asked to choose a serious game or one they had played for academic purposes; rather, they were only asked to choose a game they had previously played and with which they were familiar.

The "Alpha if item is deleted" column shows the Cronbach's alpha that would result if an item were removed from the scale. All values remain within a narrow range around the total alpha (0.8433 to 0.8615), and none of the items substantially reduce reliability. This suggests that all items contribute positively to internal consistency and should be retained.

Finally, the column labeled "Corrected Item-Total Correlation" indicates the correlation of each item with the total scale score (excluding the item itself). Values above 0.30 are generally considered acceptable. The attributes Skill-Task Balance (r = 0.6770) and Progression Structure (r = 0.6760) showed the strongest correlations, suggesting that they are highly representative of the overall construct. The lowest correlation was observed for Social Interaction (r = 0.3916), which, while acceptable, indicates less alignment with the overall scale.

4.3 Multiple regression and linear programming

To evaluate the practical relevance of the proposed instrument, two complementary analytical methods were employed to examine whether users are capable of identifying the attributes they perceive as most influential in shaping their overall assessment of a video game. To this end, the survey collected both the individual ratings users assigned to each attribute (based on their evaluation of a specific game) and their overall rating of that same game.

These data were analyzed using two techniques: (1) a multiple linear regression analysis to explore the predictive value of each attribute, and (2) the linear programming approach proposed by Pacheco-Velázquez et al. [6], which estimates the implicit weight users assign to each factor in their global judgment.

4.3.1 Multiple Linear Regression

Multiple linear regression is a statistical technique used to model the relationship between a dependent variable and two or more independent variables. In the context of this study, the primary goal is to determine whether the overall game rating can be predicted based on the individual evaluations of specific game attributes.

By using the attributes included in the survey as independent variables, the analysis seeks to identify which of them have the greatest influence on the overall score assigned by users. A stepwise regression approach was employed to construct the model iteratively, incorporating or removing variables based on their statistical significance in each step.

The results of this analysis are summarized in Table 4, while Tables 5 and 6 present the detailed outputs of the final regression model.

Table 4. Summary of the Regression Model

Metric	Value
R-squared	48.99%
Adjusted R-squared	48.49%
Standard Error of Estimate (S)	0.9686
Predicted R-squared	46.05%
F-statistic	97.96
p-value (overall model)	< 0.0001

Table 5. Results obtained by performing multiple linear regression

Term	Coefficient	SE (Coef)	T- Value	p-value	VIF
Constant	2.8060	0.2710	10.34	0	—
Concentration	0.1064	0.0276	3.85	0	1.42
Autonomy	0.0848	0.0259	3.27	0.001	1.46
Emotional Involvement	0.0941	0.0211	4.46	0	1.52
Balance Between Skills and Task	0.2156	0.0303	7.11	0	1.86
Progression Structure	0.1029	0.0275	3.74	0	1.53
Animation and Sound	0.1102	0.0276	3.99	0	1.39

Table 6. Analysis of Variance (ANOVA)

Source	DF	SS (Adj.)	MS (Adj.)	F-Value	p-value
Regression	6	551.45	91.91	97.96	0
Error	612	574.21	0.9382		
Lack of fit	554	572.73	1.0338	40.53	0
Pure error	58	1.48	0.0255		
Total	618	1125.66			

Table 4 indicates that the model explains approximately 49% of the variance in the overall game rating, which is considered acceptable given the perceptual nature of the data. Moreover, the F-value of 97.96 and a p-value < 0.0001 confirm that the model is statistically significant. The predicted R-squared value (46.05%) further suggests that the model possesses a reasonable level of generalizability.

Table 5 shows that all predictors included in the model are statistically significant ($p < 0.05$). Notably, the variable "Balance Between Skills and Tasks" stands out with the highest positive coefficient and the largest t-value, indicating it is the most influential factor in users' overall evaluation of the game. On the other hand, all the coefficients of the attributes included in the model are positive, which indicates that each of these variables has a positive effect on the overall perception of the game. Although this outcome was expected, it is worth noting that it reveals no inconsistencies when evaluating the game based on these attributes. In addition, all Variance Inflation Factor (VIF) values are below 2, ruling out concerns of multicollinearity among the predictors.

Finally, Table 6 (ANOVA) further supports the statistical validity of the overall model ($F = 97.96$, $p < 0.001$). However, the significant result in the lack-of-fit test suggests potential non-linearity or the omission of relevant variables. This opens opportunities for further research to refine and extend the current model.

4.3.2 Linear Programming

Finally, the linear programming method proposed by Pacheco-Velázquez et al. [6] was applied. This approach aims to determine the relative weight that participants (as a group) assign to each factor in explaining their overall evaluation of a video game.

By solving the linear programming model, the following results were obtained:

Table 7. Results obtained by solving the linear programming model

Attribute	Weight
Challenge	1.14%
Feedback	0.00%
Immersion	15.55%
Concentration	8.14%
Clarity of Goals	3.37%
Autonomy	7.55%
Social Interaction	0.00%
Knowledge Enhancement	0.00%
Emotional involvement	4.86%
Balance between skills and task	19.53%
Engaging Narrative	0.00%
Progression structure	10.59%
Animation and sound	29.27%

The results revealed that "Animation and Sound" received the highest weight at 29.27%, followed by "Balance between Skills and Task" with 19.53%, and "Immersion" at 15.55%. Other notable attributes included "Progression Structure" (10.59%), "Concentration" (8.14%), and "Autonomy" (7.55%). Attributes like "Challenge," "Clarity of Goals," and "Emotional Involvement" had lower weights, whereas "Feedback," "Social Interaction," "Knowledge Enhancement," and "Engaging Narrative" were assigned a weight of 0.00%. These findings highlight the key factors that players prioritize when evaluating a video game.

5. Results and Discussion

5.1 Main results

The GEMS questionnaire was administered to a total of 619 participants. The internal consistency of the scale was assessed using Cronbach's alpha, yielding a value of 0.861, which indicates a high level of reliability. This result supports the coherence of the selected attributes and confirms their relevance as distinct dimensions of the player experience.

The survey was completed primarily by high school and undergraduate students, who were allowed to freely select the video game they wished to evaluate. The only requirement was that the chosen game be one they had played regularly and felt comfortable rating in terms of the various attributes under consideration.

Under these conditions, the descriptive statistics revealed that the attributes Animation and Sound, Immersion, and Progression Structure received the highest importance ratings. In contrast, Feedback and Knowledge Enhancement were rated lower (although this does not imply that these attributes are irrelevant).

When the attributes are ranked in descending order of importance, results across both analytical methods suggest a consistent pattern. These findings reinforce the GEMS scale's utility in capturing the relative weight users assign to different game elements. A summary of this comparison is presented in Table 8.

Table 8. Summary of data obtained from the survey, multiple regression, and linear programming

Attribute	Mean perceived importance	Attribute significance in the regression model	Weight obtained using linear programming (%)
Animation and sound	8.99	Significant	29.27
Immersion	8.97	Not significant	15.55
Progression structure	8.77	Significant	10.59
Clarity of Goals	8.69	Not significant	3.37
Balance between skills and task	8.69	Very significant	19.53
Autonomy	8.67	Significant	7.55
Concentration	8.58	Significant	8.14
Challenge	8.50	Not significant	1.14
Engaging Narrative	8.43	Not significant	0.00
Emotional involvement	8.30	Significant	4.86
Social Interaction	8.01	Not significant	0.00
Feedback	7.82	Not significant	0.00
Knowledge Enhancement	7.69	Not significant	0.00

By examining Table 8, we can see that among the top seven attributes identified by users as most important, the multiple linear regression model was able to detect five of them. The attributes “Immersion” and “Clarity of Goals” were not identified by the model. Among the remaining six attributes, the regression model found only one (Emotional Involvement) to be statistically significant.

When analyzing the linear programming model, it becomes evident that all seven attributes most frequently identified by users as important were assigned positive weights in the model. Among them, “Animation and Sound” received the highest weight, surpassing all other attributes in its contribution to the overall evaluation. Notably, the combined weight of the top three attributes (“Animation and Sound”, “Immersion”, and “Progression Structure”) accounts for approximately 55% of the total weighting, suggesting that these three dimensions are central to users’ global perception of game quality. In contrast, the remaining six attributes, as ranked by users, collectively represent only about 6% of the overall weighting according to the linear programming model. It is important to note that some attributes received a weight of 0.00% in the importance analysis based on the linear programming model. This result does not imply that these attributes are irrelevant or should be removed from the GEMS framework. Rather, it reflects the outcome of the optimization process, in which certain attributes did not contribute additional explanatory power to the overall game rating once higher-weighted attributes were considered. In practical terms, a zero weight indicates that, for the analyzed sample and within the specific modeling constraints, these attributes had a negligible marginal contribution relative to other criteria. Nevertheless, these attributes were retained in the scale to preserve conceptual completeness and theoretical coherence, as their relevance may emerge in different game genres, user populations, or application contexts. Exploring such variations represents a promising direction for future research.

An interesting case is the attribute “Balance between skills and tasks”. Although it was rated as only the fifth most important attribute in users' direct responses, it emerged as the strongest predictor in the multiple regression model and was also assigned the second-highest weight in the linear programming results. Additionally, “Emotional involvement”, while not among the top-rated attributes by users, was statistically significant in the regression model and received a moderate weight (approximately 5%) in the linear programming model, further supporting its relevance in shaping overall game evaluations.

The convergence of results observed in this study supports the notion that the analytical methods employed (multiple linear regression and linear programming) are effective in identifying the attributes that carry the greatest weight in users' overall evaluation of a video game. As noted in Reference [6], decision-making tasks involving the selection or assessment of an object are inherently complex, particularly when they involve evaluating a set of attributes that are difficult to compare. The authors argue that this complexity increases substantially when the evaluation involves non-binary numerical attributes. Nevertheless, when individuals can recognize and assign a numerical value to each attribute, the evaluation process becomes significantly more manageable. Importantly, the article also highlights that people do not typically engage in conscious numerical calculations to validate the consistency of their judgments. Instead, they rely on heuristic strategies, mental shortcuts that allow them to assign value to their perceptions and make decisions more efficiently.

In this context, the alignment between the outcomes of the two analytical methods and participants' self-reported perceptions provides compelling evidence that individuals can identify and prioritizing the key factors they consider when evaluating a game.

These results suggest that low weight assigned to attributes such as Knowledge Enhancement likely reflects the fact that most participants evaluated entertainment games. As such, these findings cannot be generalized to serious games. Instead, the applicability of GEMS to educational or training contexts should be considered a promising direction for future research rather than a confirmed outcome of the present study.

5.2 Comparison with previous research

On the other hand, while the development of the GEMS scale draws on the theoretical foundations and attribute sets proposed in prior instruments (e.g., EGameFlow, GEQ, PENS, GUESS), it is important to clarify that its purpose and methodological orientation differ substantially from those of previous scales. Most existing instruments are designed to measure the degree to which specific attributes are present or fulfilled within a given game. Their primary focus is often on assessing the quality or intensity of predefined dimensions such as immersion, flow, or usability. In contrast, the present study aims to identify which attributes are perceived by users as most influential in their overall evaluation of a video game. Given this fundamental difference in objectives, we do not consider it appropriate to conduct direct comparisons between the GEMS scale and other instruments. The current approach does not seek to replicate existing measurement models, but rather to provide a complementary perspective—one that focuses on user-driven prioritization of game attributes, rather than objective fulfillment of design criteria.

This distinction underlines the unique contribution of the GEMS scale: it offers a framework for understanding how players internally weigh different factors when forming an overall judgment of a game, which may have important implications for both game design and user-centered evaluation strategies.

Compared to existing evaluation instruments, GEMS provides a distinct contribution by focusing on the relative importance of attributes in shaping players' overall judgments. For instance, the Game Experience Questionnaire (GEQ) emphasizes the intensity of emotional and experiential states, while eGameFlow evaluates aspects of engagement and flow during gameplay. The Serious Game Quality Model (SGQM), in turn, assesses the effectiveness of serious games in achieving pedagogical objectives. In contrast, GEMS does not measure how

strongly attributes are perceived, but rather how much weight players assign to them when evaluating a game holistically. This shift in focus offers a complementary perspective to established instruments and allows researchers and designers to prioritize features that matter most to users, rather than assuming equal relevance across dimensions.

5.3 Limitations

While this study offers valuable insights into players' perceptions of key video game attributes, it is important to acknowledge certain limitations.

First, the sample was composed primarily of high school and university students. Although this population is relevant for analyzing gaming habits in educational or formative contexts, it limits the generalizability of the findings to other demographic groups.

Second, participants were allowed to freely select the game they wished to evaluate. Although this approach increases ecological validity by reflecting users' authentic preferences, it also introduces variability in terms of genre, platform, complexity, and gameplay objectives. This diversity, while beneficial in some respects, makes it difficult to isolate genre-specific patterns of attribute prioritization. Subsequent studies could address this issue by incorporating stratified designs or genre-focused comparisons.

Third, although the GEMS scale was designed to be broadly applicable across gaming contexts, it may not fully capture the specific dimensions relevant to serious games—those intended for educational or training purposes. In this study, only eight participants selected a game used in an academic setting, which is insufficient for reliable conclusions about this category. As a result, the evaluation of serious games remains an open question, and future research should explicitly target this type of game to assess whether GEMS adequately captures the attributes most relevant in pedagogical or professional training contexts. Importantly, doing so would also allow for illustrative applications—for instance, determining whether attributes such as "Animation and Sound" should be prioritized by educators or designers when developing educational titles.

Fourth, the significant lack-of-fit result observed in the multiple regression analysis suggests that the relationships between variables may not be strictly linear. This indicates that unmeasured factors, interaction effects, or nonlinear patterns may influence users' evaluations and merit further exploration.

A further limitation of this study is that the validation of GEMS remains partial. While internal consistency was demonstrated through Cronbach's α, no confirmatory factor analysis (CFA) or structural equation modeling (SEM) was performed to test the dimensional structure of the scale. As GEMS was designed as a multidimensional instrument, future research should apply these techniques to examine whether the proposed structure is supported by empirical data. Such analyses would provide stronger evidence of construct validity and further reinforce the robustness of the instrument.

5.4 Future Research Directions

Building upon the current findings, several promising avenues for future research can be proposed.

First, it would be valuable to examine how the perceived importance of different attributes varies across game genres. For example, players may place greater emphasis on visual quality and sound design in sports or racing games, whereas strategic depth and clarity of objectives may be more critical in strategy or simulation games. Future studies could segment participants by genre or conduct controlled comparisons to investigate these differences.

Second, a targeted study focused on serious games is needed. Recruiting users who have interacted with games in educational, professional, or therapeutic contexts would enable researchers to explore whether their evaluation criteria differ meaningfully from entertainment-focused players. For example, if GEMS highlights “Animation and Sound” as one of the most heavily weighted attributes, designers of an educational game for medical training might prioritize realistic audiovisual simulations, as these could enhance immersion and motivation while supporting knowledge retention. Similarly, if “Progression Structure” is found to be a strong predictor, educators could adapt the reward systems and milestone tracking within a serious game to sustain engagement and encourage continuous learning.

Third, it would be relevant to analyze whether the importance assigned to attributes varies according to demographic factors, such as age, gender, educational background, or geographic region. A stratified analysis could help tailor game design or evaluation strategies to specific user segments.

Fourth, in addition to structural equation modeling, future research could explore other nonlinear approaches such as polynomial regression, which may capture curvilinear relationships between attributes and overall satisfaction, or machine learning techniques that can model complex interactions without strict parametric assumptions. These approaches could provide more nuanced insights into how game attributes combine to shape players’ holistic evaluations.

Fifth, the current findings could be enriched through qualitative methods, including interviews or open-ended questionnaires. These approaches would help uncover the reasoning behind players' preferences and offer deeper insights into the heuristics they use to evaluate games.

Sixth, employing structural equation modeling to further explore and confirm the relationships between attributes analyzed and overall user satisfaction in video games. This approach will allow for a more nuanced understanding of how attributes like "Animation and Sound," which received the highest weight of 29.27%, and "Balance between Skills and Task" at 19.53%, interact with other elements such as "Immersion," "Progression Structure," and "Concentration." Utilizing structural equation modeling could provide deeper insights into the causal pathways and interdependencies among these factors, leading to more effective game design and user-centered evaluation frameworks.

Seventh, some items in the current version of GEMS include compound formulations that combine a general construct with an explanatory or illustrative question. While this approach can be useful in certain cases, other items (such as Clarity of Objectives), may unintentionally address more than one aspect within a single question. Future versions of GEMS should refine these items by separating conceptually distinct elements or by relying on more focused formulations, in line with best practices in questionnaire design and psychometric scale development.

Eighth, user satisfaction constitutes a well-established construct with its own body of research across fields such as human–computer interaction, information systems, and game studies. In the present study, user satisfaction was operationalized as an overall evaluative outcome to examine the relative importance of game attributes. However, future research on GEMS should explicitly ground the definition of user satisfaction in the relevant literature and explore its multidimensional nature, potentially distinguishing between experiential, affective, and evaluative components.

Finally, future studies might investigate the stability of attribute prioritization over time. Do players' preferences shift as they become more experienced with a game? Are some attributes more important during early gameplay stages versus long-term engagement? These questions could be explored through longitudinal studies.

Although the present study primarily involved participants evaluating entertainment-focused games, the structure of GEMS makes it adaptable to serious games used in educational, training, or therapeutic contexts. Future applications in these domains could reveal whether players assign different levels of importance to certain attributes (e.g., knowledge enhancement, clarity of objectives) when the primary purpose of the game is learning rather than entertainment. This alignment underscores the potential of GEMS to contribute not only to commercial game evaluation but also to pedagogical and applied domains relevant to IJSG's readership.

While the proliferation of game evaluation models reflects the diversity of game types, purposes, and research traditions, it also raises questions about fragmentation and comparability across studies. Rather than advocating for the continuous creation of isolated evaluation frameworks, future research may benefit from efforts aimed at convergence and integration across complementary approaches. In this context, GEMS is not intended to replace existing models, but to contribute to a broader evaluative landscape by emphasizing the relative importance of attributes from the users' perspective. Over time, such approaches may help identify a core set of shared dimensions that could support the development of more flexible and widely applicable evaluation standards, adaptable to different game genres and application domains, including entertainment and serious games.

6. Conclusions

This study introduces GEMS (the Game Experience Multidimensional Scale), a 13-item instrument that captures the attributes players themselves consider when forming an overall judgment of a video game. Psychometric testing with 619 participants yielded a high internal reliability (Cronbach's $\alpha = 0.861$), confirming that the selected attributes constitute a coherent set of perceptual dimensions. Two complementary analytical approaches (stepwise multiple linear regression and a linear-programming weighting model) converged in identifying "Animation and Sound", "Balance between skills and tasks", "Immersion", and "Progression Structure" as the most influential predictors of global game ratings, together explaining roughly half of the variance and accounting for approximately 75 % of the total attribute weight.

These findings demonstrate that players can discern and prioritize the factors they deem critical to game quality, and that GEMS can quantify those implicit weightings. Beyond confirming the relevance of flow-related constructs such as balance and immersion, the results highlight the growing importance of audiovisual aesthetics and reward structures, elements that may be under-represented in traditional flow or motivation-centric scales. By shifting the focus

from checking whether design criteria are present to understanding which criteria matter most to users, GEMS offers a practical diagnostic tool for designers, educators, and researchers seeking to create or evaluate games that resonate more deeply with their target audiences. A key next step is to apply GEMS exclusively to serious games in order to test whether the relative importance of attributes differs when games are designed for learning, professional development, or therapeutic purposes. Establishing these differences will be crucial for positioning GEMS as a robust tool for serious-game evaluation.

In summary, GEMS should be regarded as a preliminary but promising instrument for examining how players prioritize different game attributes when forming overall evaluations. The present study provides initial evidence of reliability and analytical value, yet further work is needed to expand its validation, diversify the participant base, and apply the instrument specifically to serious games. By clarifying its scope and acknowledging its current limitations, GEMS lays the groundwork for a more comprehensive understanding of attribute prioritization in game evaluation and has the potential to evolve into a significant contribution to both research and practice.

Conflicts of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

[1] S. Egenfeldt-Nielsen, J. H. Smith, and S. P. Tosca, *Understanding video games: the essential introduction*, Fourth edition. New York London: Routledge, Taylor & Francis Group, 2020. doi: 10.4324/9780429431791.

[2] M. De Aguilera and A. Mendiz, "Video games and education: (Education in the Face of a 'Parallel School')," *Comput. Entertain.*, vol. 1, no. 1, pp. 1–10, Oct. 2003, doi: 10.1145/950566.950583.

[3] T. L. Taylor, *Watch Me Play: Twitch and the Rise of Game Live Streaming.* Princeton University Press., 2018.

[4] M. R. Johnson and J. Woodcock, "Work, play, and precariousness: An overview of the labour ecosystem of esports," *Media, Culture & Society*, vol. 43, no. 8, pp. 1449–1465, Nov. 2021, doi: 10.1177/01634437211011555.

[5] M. Sjöblom and J. Hamari, "Why do people watch others play video games? An empirical study on the motivations of Twitch users," *Computers in Human Behavior*, vol. 75, pp. 985–996, Oct. 2017, doi: 10.1016/j.chb.2016.10.019.

[6] E. Pacheco-Velazquez, V. Rodes-Paragarino, L. Rabago-Mayer, and A. Bester, "A new methodology to determine the importance of game attributes.," *IJSG*, vol. 11, no. 4, pp. 41–56, Nov. 2024, doi: 10.17083/ijsg.v11i3.802.

[7] F.-L. Fu, R.-C. Su, and S.-C. Yu, "EGameFlow: A scale to measure learners' enjoyment of e-learning games," *Computers & Education*, vol. 52, no. 1, pp. 101–112, Jan. 2009, doi: 10.1016/j.compedu.2008.07.004.

[8] S. De Freitas and M. Oliver, "How can exploratory learning with games and simulations within the curriculum be most effectively evaluated?," *Computers & Education*, vol. 46, no. 3, pp. 249–264, Apr. 2006, doi: 10.1016/j.compedu.2005.11.007.

[9] J. H. Brockmyer, C. M. Fox, K. A. Curtiss, E. McBroom, K. M. Burkhart, and J. N. Pidruzny, "The development of the Game Engagement Questionnaire: A measure of engagement in video game-playing," *Journal of Experimental Social Psychology*, vol. 45, no. 4, pp. 624–634, July 2009, doi: 10.1016/j.jesp.2009.02.016.

[10] L. del C. García-Mundo, "QSGAME-MODEL: A serious game product quality model," PhD Thesis, Results obtained by performing multiple linear regression, Spain, 2016.
[11] K. Mitgutsch and N. Alvarado, "Purposeful by design?: a serious game design assessment framework," in *Proceedings of the International Conference on the Foundations of Digital Games*, Raleigh North Carolina: ACM, May 2012, pp. 121–128. doi: 10.1145/2282338.2282364.
[12] N. D. Bowman, "Editorial: Video Games as Demanding Technologies," *MaC*, vol. 7, no. 4, pp. 144–148, Dec. 2019, doi: 10.17645/mac.v7i4.2684.
[13] S. Deterding, D. Dixon, R. Khaled, and L. Nacke, "From game design elements to gamefulness: defining 'gamification,'" in *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, Tampere Finland: ACM, Sept. 2011, pp. 9–15. doi: 10.1145/2181037.2181040.
[14] D. R. Michael and S. L. Chen, *Serious games: Games that educate, train, and inform.*, Muska&Lipman/Premier-Trade. 2005.
[15] F. Bellotti, B. Kapralos, K. Lee, P. Moreno-Ger, and R. Berta, "Assessment in and of Serious Games: An Overview," *Advances in Human-Computer Interaction*, vol. 2013, pp. 1–11, 2013, doi: 10.1155/2013/136864.
[16] C. C. Abt, *Serious games.* University press of America., 1987.
[17] P. Backlund and M. Hendrix, "Educational games - Are they worth the effort? A literature survey of the effectiveness of serious games," in *2013 5th International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES)*, Poole: IEEE, Sept. 2013, pp. 1–8. doi: 10.1109/VS-GAMES.2013.6624226.
[18] E. L.-C. Law and X. Sun, "Evaluating user experience of adaptive digital educational games with Activity Theory," *International Journal of Human-Computer Studies*, vol. 70, no. 7, pp. 478–497, July 2012, doi: 10.1016/j.ijhcs.2012.01.007.
[19] J. Moizer *et al.*, "An approach to evaluating the user experience of serious games," *Computers & Education*, vol. 136, pp. 141–151, July 2019, doi: 10.1016/j.compedu.2019.04.006.
[20] M. H. Phan, J. R. Keebler, and B. S. Chaparro, "The Development and Validation of the Game User Experience Satisfaction Scale (GUESS)," *Hum Factors*, vol. 58, no. 8, pp. 1217–1247, Dec. 2016, doi: 10.1177/0018720816669646.
[21] C. Jennett *et al.*, "Measuring and defining the experience of immersion in games," *International Journal of Human-Computer Studies*, vol. 66, no. 9, pp. 641–661, Sept. 2008, doi: 10.1016/j.ijhcs.2008.04.004.
[22] B. G. Witmer and M. J. Singer, "Measuring Presence in Virtual Environments: A Presence Questionnaire," *Presence*, vol. 7, no. 3, pp. 225–240, June 1998, doi: 10.1162/105474698565686.
[23] S. A. Jackson and H. W. Marsh, "Development and Validation of a Scale to Measure Optimal Experience: The Flow State Scale," *Journal of Sport and Exercise Psychology*, vol. 18, no. 1, pp. 17–35, Mar. 1996, doi: 10.1123/jsep.18.1.17.
[24] S. A. Jackson and R. C. Eklund, "Assessing Flow in Physical Activity: The Flow State Scale–2 and Dispositional Flow Scale–2," *Journal of Sport and Exercise Psychology*, vol. 24, no. 2, pp. 133–150, June 2002, doi: 10.1123/jsep.24.2.133.
[25] R. M. Ryan, C. S. Rigby, and A. Przybylski, "The Motivational Pull of Video Games: A Self-Determination Theory Approach," *Motiv Emot*, vol. 30, no. 4, pp. 344–360, Dec. 2006, doi: 10.1007/s11031-006-9051-8.
[26] E. L. Deci and R. M. Ryan, *Intrinsic Motivation and Self-Determination in Human Behavior*. Boston, MA: Springer US, 1985. doi: 10.1007/978-1-4899-2271-7.
[27] G. F. Tondello, R. R. Wehbe, L. Diamond, M. Busch, A. Marczewski, and L. E. Nacke, "The Gamification User Types Hexad Scale," in *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*, Austin Texas USA: ACM, Oct. 2016, pp. 229–243. doi: 10.1145/2967934.2968082.
[28] O. Schaffer, "Development and Preliminary Validation of the Enjoyment Questionnaire and the Sources of Enjoyment Questionnaire," in *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, New Orleans LA USA: ACM, Apr. 2022, pp. 1–7. doi: 10.1145/3491101.3519819.

[29] E. Fokides, P. Atsikpasi, P. Kaimara, and I. Deliyannis, “Let players evaluate serious games. Design and validation of the Serious Games Evaluation Scale,” *ICG*, vol. 41, no. 3, pp. 116–137, Nov. 2019, doi: 10.3233/ICG-190111.
[30] F. Zurita Ortega, N. Medina Medina, F. L. Gutiérrez Vela, and R. Chacón Cuberos, “Validation and Psychometric Properties of the Gameplay-Scale for Educative Video Games in Spanish Children,” *Sustainability*, vol. 12, no. 6, p. 2283, Mar. 2020, doi: 10.3390/su12062283.
[31] K. Martinez, M. I. Menéndez-Menéndez, and A. Bustillo, “A New Measure for Serious Games Evaluation: Gaming Educational Balanced (GEB) Model,” *Applied Sciences*, vol. 12, no. 22, p. 11757, Nov. 2022, doi: 10.3390/app122211757.
[32] E. Oprins, G. Visschedijk, M. B. Roozeboom, M. Dankbaar, W. Trooster, and S. C. E. Schuit, “The game-based learning evaluation model (GEM): measuring the effectiveness of serious games using a standardised method,” *IJTEL*, vol. 7, no. 4, p. 326, 2015, doi: 10.1504/IJTEL.2015.074189.
[33] J. L. González Sánchez, “Jugabilidad. Caracterización de la experiencia del jugador en videojuegos,” Tesis Doctoral, Universidad de Granada, España, 2010.
[34] J. L. González Sánchez, N. P. Zea, F. L. Gutiérrez, and M. J. Cabrera, “De la usabilidad a la jugabilidad: Diseño de videojuegos centrado en el jugador.,” presented at the IX Congreso Internacional Interacción, Albacete, España, June 2008.
[35] M. Carrión-Toro *et al.*, “Rethinking Usability in Serious Games: Designing an Instrument That Evaluates What Really Matters in Learning Contexts,” *Applied Sciences*, vol. 15, no. 18, p. 10084, Sept. 2025, doi: 10.3390/app151810084.
[36] C. S. Fernandes, M. J. Campos, M. T. Moreira, A. Lima, S. Ferreira, and M. M. Martins, “Development and Validation of the Serious Educational Game in Nursing Appraisal Scale,” *Nursing Reports*, vol. 14, no. 2, pp. 1148–1157, May 2024, doi: 10.3390/nursrep14020087.
[37] L. Bunt, J. Greeff, and E. Taylor, “Enhancing Serious Game Design: Expert-Reviewed, Stakeholder-Centered Framework,” *JMIR Serious Games*, vol. 12, p. e48099, May 2024, doi: 10.2196/48099.
[38] M. Zeiler *et al.*, “Game Design, Effectiveness, and Implementation of Serious Games Promoting Aspects of Mental Health Literacy Among Children and Adolescents: Systematic Review,” *JMIR Ment Health*, vol. 12, p. e67418, May 2025, doi: 10.2196/67418.
[39] L. A. Clark and D. Watson, “Constructing validity: Basic issues in objective scale development.,” *Psychological Assessment*, vol. 7, no. 3, pp. 309–319, Sept. 1995, doi: 10.1037/1040-3590.7.3.309.
[40] A. Stefana *et al.*, “Psychological, psychiatric, and behavioral sciences measurement scales: best practice guidelines for their development and validation,” *Front. Psychol.*, vol. 15, p. 1494261, Jan. 2025, doi: 10.3389/fpsyg.2024.1494261.